chinaXiv:201911.00093v1

社会困境研究的范式沿革及其理论价值：以道德两难决策为例

刘传军　廖江群*

（清华大学社会科学学院心理学系, 北京 100084）

摘要：社会困境是伦理学、社会学、心理学和经济学等多学科的共同研究主题，以道德两难决策为代表。在综合分析道德决策研究中的三种实证研究范式（经典两难法、加工分离法和 CNI 模型法）沿革过程的基础上，本文提出了 CAN 维度合成法，解决了前述研究范式的理论和方法的局限性。这四种方法的发展，使道德困境研究不再局限于冲突性的两难情境探讨，从而拓展了道德情境研究的范畴，推动了对现有道德理论的实证检验和对研究争议的解决，并为相关学科中具有潜在冲突性的研究议题提供了方法论借鉴。

关键词：社会困境；道德两难；规则冲突；范式沿革；实证方法

一、前言

随着无人驾驶汽车的到来，社会困境问题再次引起热议。与之前那些看上去在现实中发生的可能性并不高的假定情境相比，无人驾驶汽车的到来，将困境问题实实在在地呈现在人们面前。①当无人驾驶汽车在行驶过程中遇到紧急情况时，如果按原定路线直接开过去将撞死 5 个甚至更多行人，如果通过改变方向撞向路牙则会牺牲汽车及其乘客。在这种紧急情况之下，无人驾驶汽车应该如何反应？无人驾驶汽车背后的程序设置本质上反映了人们的抉择倾向。近期研究发现，人们倾向于牺牲汽车及其乘客但并不愿意乘坐或购买这样的汽车。②③

与无人驾驶汽车类似的社会两难困境，最早出现在道德领域，被称为道德两难情境，而后扩展到社会生活的诸多方面。道德困境研究的兴起具有实证主义和常人道德的大背景。在道德困境研究范式形成之前，道德研究一直使用的是哲学思辨方法。这种方法至少具有三方面局限：一是各个流派哲学家的基本立场不同，从而对道德的基本观点本身就存在差异，导致争论难以达成一致；二是哲学家思辨下的道德与普通人眼中的道德未必一致，而常人视角下的道德更具现实意义和社会意义；三是这些道德观点停留在哲学思辨上而具有不可检验性，不能从客观上对其直接检验。在这样的背景之下，道德研究亟需发展出直观、可检验并能被普通人所理解的研究范式，道德困境的思想实验也就应运而生。

---

* 本研究为国家社会科学基金项目(18BSH114)和清华大学自主科研计划项目(2017THZWYY11)的阶段性成果。

通讯作者：廖江群，E-mail: liaojq@tsinghua.edu.cn.

① A. D. Young and A. E. Monroe, “Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas,” *Journal of Experimental Social Psychology*, vol. 85, 2019: 103870.

② J. F. Bonnefon, A. Shariff, and I. Rahwan, “The social dilemma of autonomous vehicles,” *Science*, vol. 352, no. 6293, 2016, pp. 1573-1576.

③ J. D. Greene, “Our driverless dilemma When should your car be willing to kill you?” *Science*, vol. 352, no. 6293, 2016, pp. 1514-1515.

1967 年，福特提出了著名的电车难题，标志着道德困境研究范式的形成。①大约十年后，汤姆森又提出了另一个经典道德困境——天桥范式。②电车难题中，决策者需要通过选择是否扳道岔或者按开关来决定是否牺牲一个无辜者以拯救五个人，而天桥范式中，决策者需要选择是否推倒一个体型硕大的陌生人来决定是否牺牲这个人以拯救其他五个人。对于牺牲一人（或少数人）而拯救五人（或多数人）的提议，表示赞同则被认为是功利主义的道德判断，因为这种判断是基于最大化结果来考虑的，符合最大化人类福祉的功利主义原则；而表示拒绝则被认为是义务论的道德判断，因为这种判断被认为是基于行为本身是否符合道德规范而来的，符合义务论原则对行为本身的规范合理性的要求。

电车难题的行为实现方式比较间接，无辜者的死亡是由电车直接导致而非决策者亲手造成，因此被称为非个人道德困境(impersonal dilemma)。相反，天桥范式中，无辜者的死亡是由于决策者直接推倒从而亲手造成其死亡，因此被称为个人道德困境(personal dilemma)。这两类困境在后续的道德研究中被奉为经典，许多研究探讨了这两种困境及其变型中人们的决策偏好及其成因与后果。③④⑤⑥比如有研究者发现男性在个人道德两难中更加功利主义，但在非个人道德范式中无性别差异；⑦语言描述框架也可能会产生作用，比如行为提议描述为“救命(save lives)”或者“不要杀人(do not kill)”，人们的决策倾向也会有所不同，⑧具体可见徐同洁等人的评述。⑨道德两难的应用非常广泛，在我国的道德教育中也常常使用道德两难故事法来进行课堂讨论，从而提升学生的道德认知水平。⑩⑪

随着研究推进，许多道德两难困境及其变型被开发出来，研究者所使用的情境材料各异，缺乏标准化的研究方法，导致许多元分析因缺乏研究方法的共通性而无法进行。

---

① P. Foot, “The Problem of Abortion and the Doctrine of Double Effect,” *Oxford Review*, vol. 2, no. 2, 1967, pp. 152–161.
② J. J. Thomson, “Killing, letting die, and the trolley problem,” *Monist*, vol. 59, no. 2, 1976, pp. 204-217.
③ J. D. Greene, R. B. Sommerville, L. E. Nystrom, et al., “An fMRI investigation of emotional engagement in moral judgment,” *Science*, vol. 293, no. 5537, 2001, pp. 2105-2108.
④ P. Valdesolo and D. Desteno, “Manipulations of emotional context shape moral judgment,” *Psychological Science*, vol. 17, no. 6, 2006, pp. 476-477.
⑤ N. Gold, A. M. Colman and B. D. Pulford, “Cultural differences in responses to real-life and hypothetical trolley problems,” *Judgment and Decision Making*, vol. 9, no. 1, 2014, pp. 65-76.
⑥ J. Graham, P. Meindl, E. Beall, et al., “Cultural differences in moral judgment and behavior, across and within societies,” *Current Opinion in Psychology*, vol. 8, 2016, pp. 125-130.
⑦ M. Fumagalli, R. Ferrucci, F. Mameli, et al., “Gender-related differences in moral judgments,” *Cognitive Process*, vol. 11, no. 3, 2010, pp. 219-226.
⑧ R. Broeders, K. van den Bos, P. A. Muller, et al., “Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible,” *Journal of Experimental Social Psychology*, vol. 47, no. 5, 2011, pp. 923-934.
⑨ 徐同洁,刘燕君,胡平,彭申立：《道德两难困境范式在心理学研究中的使用:回顾与展望》,《中国临床心理学杂志》2018 年第 26 卷第 3 期.
⑩ 刘锋,刘力平,吴冈,李淑敏,张毅洁：《提高道德判断能力的好方法——道德两难故事课堂讨论》,《道德与文明》1993 年第 6 期.
⑪ 肖静,潘泽江：《论"道德两难"在德育教学中的运用》,《中南民族大学学报：人文社会科学版》2005 年第 25 卷第 5 期.

chinaXiv:201911.00093v1

①此外，有研究指出，情境材料的字数，表达方式和问题形式都会影响道德判断，在研究中均需要进行标准化。⑪而在材料使用背后，更为重要的是这些道德困境研究的实证方法本身存在着差异性，厘清道德困境研究中的方法论范式对于道德困境研究以及类似的社会困境的研究，均具有重要意义。本文将对目前主流使用的经典两难法、加工分离法、CNI(Consequence, Norm and generalized Inaction preference, CNI)模型法进行综合分析，在此基础上提出 CAN(Consequence sensitivity, generalized Action preference and Norm sensitivity, CAN)维度合成法并论述其背后的方法论价值。

二、现有三种困境研究范式

（一）经典两难法

经典两难法主要使用电车难题、天桥范式等情境故事及其变型，让人们对这种假定故事中行为提议的道德正当性及其行为意愿进行评估，以此来探究人们的道德倾向。在假定情境中，需要牺牲一个无辜者（或少数人）来拯救五个人（或多数人），决策者被提问这样做在道德上是否可接受以及是否愿意执行这种行为提议等。比如在天桥范式中，决策者首先会阅读到：

"一辆失控的电车正向铁轨上的五名工人冲去，如果电车继续前进的话，他们将被杀死。你正在轨道上方的人行天桥上，天桥处于飞驰的电车和五个工人之间。在这条人行天桥旁边，有一个身材巨大的陌生人。救那五个工人生命的唯一办法就是把这个陌生人从桥上推到铁轨下面，他巨大的身躯将使电车停下来。如果你这样做，那名陌生人会死，但五个工人将会得救。"然后，决策者被提问"你认为'把陌生人推倒以挽救那五个工人'的提议在道德上是否可接受？"或者"你是否会把陌生人推倒以挽救那五个工人？"

道德双加工理论就是使用经典道德两难法发展而成的。此前的道德理论多以理性认知为基础，以皮亚杰③、科尔伯格④等为代表，认为理性推理是道德判断的基础。而 Greene 发现情绪卷入的变化会影响道德判断，并使用功利核磁共振方法初步对道德判断的神经机制进行了讨论。在 Greene 早期的研究中，与道德哲学家 Haidt 有过许多合作，强调情绪的道德效用，并通过实验证明在理性思考之外，情绪在影响道德判断方面起着重要作用，特别是对个体的义务论倾向具有预测力。⑤⑥而后来，Greene 也发现理

① J. F. Christensen and A. Gomila, "Moral dilemmas in cognitive neuroscience of moral decision-making: a principled review," *Neuroscience and Biobehavioral Reviews*, vol. 36, no. 4, 2012, pp. 1249-1264.
② J. F. Christensen, A. Flexas, M. Calabrese, et al., "Moral judgment reloaded: a moral dilemma validation study," *Frontiers in Psychology*, vol. 5, 2014: 607.
③ J. Piaget, *The moral judgment of the child*, New York, NY, US: Free Press, 1948.
④ L. Kohlberg, "Stage and sequence: The cognitive developmental approach to socialization," in D. A. GOSLING S. eds., *Handbook of socialization theory and research*. Chicago: Rand McNally, 1969, pp. 347-480.
⑤ J. D. Greene, R. B. Sommerville, L. E. Nystrom, et al., "An fMRI investigation of emotional engagement in moral judgment," *Science*, vol. 293, no. 5537, 2001, pp. 2105-2108.
⑥ J. D. Greene and J. Haidt, "How (and where) does moral judgment work?" *Trends in Cognitive Sciences*, vol. 6, no. 12, 2002, pp. 517-523.

chinaXiv:201911.00093v1

性思考对道德判断同样起着非常重要的作用，特别是对功利主义倾向具有预测力。①②由此，Greene 提出道德双加工理论，③对道德研究起着深远影响，后续许多研究都在该理论框架下展开。④⑤⑥关于该理论的评述可参见喻丰等人的分析。⑦

使用经典两难法以及基于此发展出的道德双加工理论，在后续研究中发现了许多不合常理或者分歧甚至矛盾的结果。首先，在功利主义反应倾向方面，有研究发现功利主义反应与决策者的精神病性、马基雅维里主义和生活无意义感存在着正相关，⑧特别是精神病性与功利主义反应的正相关，在后续研究中多次被证实。⑨⑩⑪对行为结果的利弊权衡一直被认为是功利主义理性的表现，并且道德双加工理论认为理性加工会驱动功利主义反应，而这却与精神病性具有正相关关系，这意味着精神病性越强或者精神病人相较于常人，更加理性和关心最大化人类福祉，或者更加理性和关心最大化人类福祉的人可能会有更强的精神病性倾向，这在理论和常识上都是不可思议的。

其次，义务论反应倾向方面，厌恶与义务论倾向的正相关关系也被许多研究探讨，⑫⑬多认为厌恶刺激唤起了决策者对违背道德规范的行为的厌恶感，特别是对纯洁性的违背具有更强厌恶感，从而驱动了更严厉的道德标准，做出更义务论的道德判断。

chinaXiv:201911.00093v1

---

① J. D. Greene, S. A. Morelli, K. Lowenberg, et al., "Cognitive load selectively interferes with utilitarian moral judgment," *Cognition*, vol. 107, no. 3, 2008, pp. 1144-1154.
② J. D. Greene, "Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains," *Trends in Cognitive Science*, vol. 11, no. 8, 2007, pp. 322-323; author reply 3-4.
③ J. D. Greene, "Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie," *Journal of Experimental Social Psychology*, vol. 45, no. 3, 2009, pp. 581-584.
④ F. F. Youssef, K. Dookeeram, V. Basdeo, et al., "Stress alters personal moral decision making," *Psychoneuroendocrinology*, vol. 37, no. 4, 2012, pp. 491-498.
⑤ E. Gubbins and R. M. J. Byrne, "Dual processes of emotion and reason in judgments about moral dilemmas," *Thinking & Reasoning*, vol. 20, no. 2, 2014, pp. 245-268.
⑥ A. Skulmowski, A. Bunge, K. Kaspar, et al., "Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study," *Frontiers in Behavioral Neuroscience*, vol. 8, 2014: 426.
⑦ 喻丰,彭凯平,韩婷婷,柴方圆,柏阳：《道德困境之困境——情与理的辩争》,《心理科学进展》2011 年第 11 期.
⑧ D. M. Bartels and D. A. Pizarro, "The mismeasure of morals: antisocial personality traits predict utilitarian responses to moral dilemmas," *Cognition*, vol. 121, no. 1, 2011, pp. 154-161.
⑨ A. Seara-Cardoso, H. Dolberg, C. Neumann, et al., "Empathy, morality and psychopathic traits in women," *Personality and Individual Differences*, vol. 55, no. 3, 2013, pp. 328-333.
⑩ S. Tassy, C. Deruelle, J. Mancini, et al., "High levels of psychopathic traits alters moral choice but not moral judgment," *Frontiers in Human Neuroscience*, vol. 7, no. 4, 2013: 229.
⑪ G. Yu and S. Tang, "Psychopathic personality and utilitarian moral judgment in college students," *Journal of Criminal Justice*, vol. 41, no. 5, 2013, pp. 342-349.
⑫ D. Pizarro, Y. Inbar and C. Helion, "On Disgust and Moral Judgment," *Emotion Review*, vol. 3, no. 3, 2011, pp. 267-268.
⑬ H. A. Chapman and A. K. Anderson, "Things rank and gross in nature: a review and synthesis of moral disgust," *Psychological Bulletin*, vol. 139, no. 2, 2013, pp. 300-327.

①②③另一方面，积极情绪则可以抵消一定厌恶感，从而减少义务论反应而增加功利主义反应，④但也有研究指出快乐(Mirth)与升华感(Elevation)两种积极情绪，其作用完全不同，前者会降低义务论倾向而后者则恰恰相反。⑤因此，厌恶情绪与义务论倾向的关系可能更加复杂，在经典两难法的方法论框架下难以得到合理解释。

再次，直觉/理性加工与义务论/功利主义反应是否是严格对应的关系？从心理模型理论角度，直觉功利主义是可能存在的，⑥而近期研究使用二次回答范式发现个体在直觉上就可能是功利主义的，⑦⑧个体在直觉加工状态下进行判断和在理性加工状态下进行再次判断，其功利主义判断结果总体上是没有变化的。并且，认知反省测试与功利主义反应之间并不具有显著的相关关系。⑨因此，直觉/理性加工与义务论/功利主义反应之间的对应关系也需要重新检验。甚至，有批评者认为在两难框架下，所谓的功利主义道德判断并不能反映人们对最大化结果的关注，并进一步质疑了经典两难法在揭示人们的道德抉择方面的有效性。⑩

经典两难法的最大局限在于它将义务论和功利主义原则严格对立起来，并且在操作上将决策者接受功利性提议与受到功利主义原则驱动划等号，而将决策者不接受功利性提议与受到义务论原则驱动划等号。这样操作造成的第一个局限是功利主义和义务论倾向是不相容的，功利主义倾向越强，义务论倾向就会越弱，反之亦然。这与人们的常识不符，也与最近的研究发现相违背。常识上，面对一个道德选择时，人们完全有可能同时综合考虑行为背后的道德规范以及行为将会造成的道德后果。近期神经研究证据表明，前扣带回、脑岛和颞上回与情感评估相关，而颞顶联合区和背内侧前额叶皮质与功利评估相关，而整体道德价值判断则表征在腹内侧前额皮质的前部；至关重要的是，这三组区域之间的反应和功能互动模式表明，情感和功利评估是独立并行计算的，并传递到腹

chinaXiv:201911.00093v1

---

① J. Haidt, P. Rozin, C. Mccauley, et al., "Body, psyche, and culture: The relationship between disgust and morality," *Psychology & Developing Societies*, vol. 9, no. 9, 1997, pp. 107-131.
② P. Rozin, L. Lowery, S. Imada, et al., "The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity)," *Journal of Personality and Social Psychology*, vol. 76, no. 4, 1999, pp. 574-586.
③ F. M. A. Wagemans, M. J. Brandt and M. Zeelenberg, "Disgust sensitivity is primarily associated with purity-based moral judgments," *Emotion*, vol. 18, no. 2, 2018, pp. 277-289.
④ P. Valdesolo and D. Desteno, "Manipulations of emotional context shape moral judgment," *Psychological Science*, vol. 17, no. 6, 2006, pp. 476-477.
⑤ N. Strohminger, R. L. Lewis and D. E. Meyer, "Divergent effects of different positive emotions on moral judgment," *Cognition*, vol. 119, no. 2, 2011, pp. 295-300.
⑥ M. Bucciarelli, "Moral dilemmas in females: children are more utilitarian than adults," *Frontiers in Psychology*, vol. 6, 2015: 1345.
⑦ B. Bago and W. De Neys, "Advancing the specification of dual process models of higher cognition: a critical test of the hybrid model view," *Thinking & Reasoning*, 2019, DOI: 10.1080/13546783.2018.1552194.
⑧ B. Bago and W. De Neys, "The intuitive greater good: Testing the corrective dual process model of moral cognition," *Journal of Experimental Psychology: General*, vol. 148, no. 10, 2019, pp. 1782-1801.
⑨ E. B. Royzman, J. F. Landy and R. F. Leeman, "Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context," *Cognitive Science*, vol. 39, no. 2, 2015, pp. 325-352.
⑩ G. Kahane, J. A. Everett, B. D. Earp, et al., "'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good," *Cognition*, vol. 134, 2015, pp. 193-209.

内侧前额叶皮质，在那里，它们被整合成一个整体的道德价值判断。①这意味着，功利主义和义务论反应倾向完全可以是相互独立，并行不悖的。

经典两难法的第二个局限在于无法量化出决策者在功利主义和义务论反应倾向上的程度差异。当这两种倾向程度相差很大时，决策者立马就可以做出判断，而当这两种倾向程度相差很小时，决策者需要更多的时间，但可能也做出了相同的判断。特别是在二元选择的情况下，决策要么接受要么不接受功利性提议，这种反应倾向的程度差异无法得到体现。②

经典两难法的第三个局限在于其解释的模糊性。③该方法将功利主义原则与义务论原则严格对立，功利主义反应增多时（比如在类似电车难题中，选择按下开关的概率增加时）既可以被认为是功利主义倾向得到强化的结果，也可以被认为是义务论倾向被弱化的结果，因此，在解释上存在着模糊性。此外，决策者赞成功利性提议，既可能是功利主义原则驱动的结果，也可能只是一般性接受提议偏好的结果（比如电车难题中，决策者根本不管是否符合道德规范，也不管是否结果有利，只是一般性地偏好接受按下开关的提议）。同理，拒绝功利性提议，既可能是义务论原则驱动的结果，也可能是一般性不接受偏好的结果。

为了解决前两个局限，研究者提出了加工分离法(Process Dissociation, PD)，而为了解决前述所有局限，研究者发展出了 CNI 模型法，以及本文提出的 CAN 维度合成法。

（二）加工分离法

加工分离法是 Conway 和 Gawronski 由记忆研究中的加工分离程序④(process dissociation procedure, PD)发展而来。Jacoby 最早使用 PD 方法来分离记忆表现当中的回忆成分和基于熟悉性猜测的成分，而这种方法并非内容特异性的，可以应用在许多领域当中。⑤Conway 和 Gawronski 第一次将它引入了道德心理学研究，以区分功利主义原则和义务论原则的驱动程度。⑥

加工分离法实质上是操控了行为提议所造成结果的利弊情况，当利大于弊时与伤害性提议构成不一致情境，即虽然行为结果有利但行为本身不符合道德规范，功利主义原则要求决策者接受该提议而义务论原则要求决策者拒绝该提议；而当弊大于利时则与伤害性提议构成一致情境，即行为结果有弊并且行为本身不符合道德规范，功利主义原则

① C. A. Hutcherson, L. Montaser-Kouhsari, J. Woodward, et al., “Emotional and Utilitarian Appraisals of Moral Dilemmas Are Encoded in Separate Areas and Integrated in Ventromedial Prefrontal Cortex,” *The Journal of Neuroscience*, vol. 35, no. 36, 2015, pp. 12593-12605.
② P. Conway and B. Gawronski, “Deontological and utilitarian inclinations in moral decision making: a process dissociation approach,” *Journal of Personality and Social Psychology*, vol. 104, no. 2, 2013, pp. 216-235.
③ B. Gawronski and J. S. Beer, “What makes moral dilemma judgments "utilitarian" or "deontological"?” *Social Neuroscience*, vol. 12, no. 6, 2017, pp. 626-632.
④ L. L. Jacoby, “A process dissociation framework: Separating automatic from intentional uses of memory,” *Journal of Memory and Language*, vol. 30, no. 5, 1991, pp. 513-541.
⑤ R. Friesdorf, P. Conway and B. Gawronski, “Gender differences in responses to moral dilemmas: a process dissociation analysis,” *Personality and Social Psychology Bulletin*, vol. 41, no. 5, 2015, pp. 696-713.
⑥ P. Conway and B. Gawronski, “Deontological and utilitarian inclinations in moral decision making: a process dissociation approach,” *Journal of Personality and Social Psychology*, vol. 104, no. 2, 2013, pp. 216-235.

chinaXiv:201911.00093v1

和义务论原则均要求决策者拒绝该提议。其原理如图 1 所示，不一致情境等同于经典两难情境，如前述的电车难题和天桥范式；一致情境将同一情境中行为提议所造成结果调整为弊大于利，比如电车难题中，失控的电车在原来轨道上行进只会轧死 1 名工人，若按下开关使其转入另一车道则会轧死 5 名工人，此时无论是受功利主义原则还是义务论原则驱动，这种伤害都是不可接受的；但若二者均未驱动，则人们可能会认为这种伤害也是可接受的。

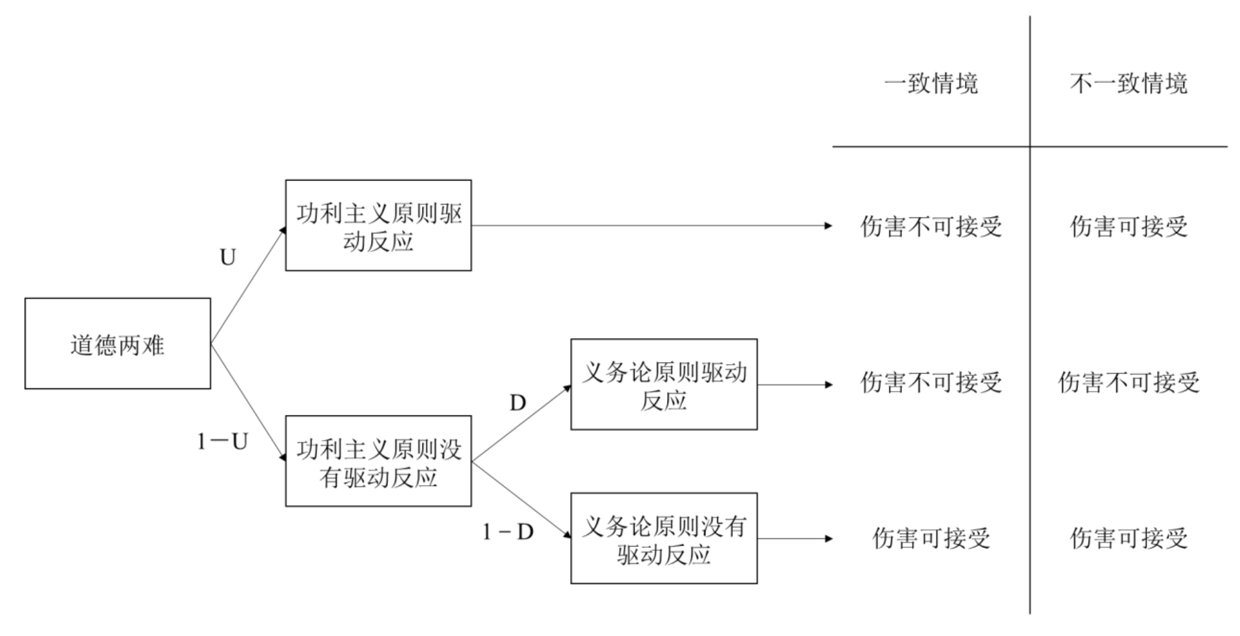


图 1 加工分离法所使用的决策加工树图解，其中 U 为 Utilitarianism 的缩写，代表功利主义原则驱动程度；D 为 Deontology 的缩写，代表义务论原则的驱动程度。①

在计算原理上，加工分离法使用了多组情境，每组情境均通过调整行为结果的利弊关系形成一致和不一致情境两个变式，以决策加工树的形式呈现出来。然后，通过计算决策者在一致情境和不一致情境中的选择伤害可接受或不可接受的概率来计算其功利主义和义务论倾向程度。因为一致情境和不一致情境条件下均有多个情境故事，因此，可以分别计算出决策者在两种条件下选择接受与不接受的反应概率。然后，使用这种反应概率推算出功利主义原则驱动程度与义务论原则驱动程度，具体计算如下：

$$p(\text{伤害不可接受}|\text{一致情境}) = U + (1-U) \times D \quad (1)$$

$$p(\text{伤害可接受}|\text{一致情境}) = (1-U) \times (1-D) \quad (2)$$

$$p(\text{伤害不可接受}|\text{不一致情境}) = (1-U) \times D \quad (3)$$

$$p(\text{伤害可接受}|\text{不一致情境}) = U + (1-U) \times (1-D) \quad (4)$$

(1)~(4)等式左侧是决策者在一致和不一致情境下接受和不接受行为提议的概率，可以从决策者的决策结果直接计算出来，比如如果在不一致情境下有 8 个情境故事，决策者在这 8 个故事中的 6 个故事选择了接受，那么，p(伤害可接受|不一致情境) = 0.75，而 p(伤害不可接受|不一致情境) = 0.25。同理，也可以计算出一致情境下决策者接受和不接受行为提议的概率。然后使(1)~(4)换算出 U 和 D 的值，比如，通过等式(4)减去等

① P. Conway and B. Gawronski, “Deontological and utilitarian inclinations in moral decision making: a process dissociation approach,” *Journal of Personality and Social Psychology*, vol. 104, no. 2, 2013, pp. 216-235.

chinaXiv:201911.00093v1

式(2)，或者通过等式(1)减去等式(3)可以得到：

U = p(伤害可接受|不一致情境)−p(伤害可接受|一致情境)

或

U = p(伤害不可接受|一致情境)−p(伤害不可接受|不一致情境)

计算出 U 的值之后，再结合等式(1)~(4)当中的任何一个等式即可计算出 D 的值。

加工分离法相对于经典两难法，在理论和方法上均往前推了一步，在经典两难法的单类型情境基础上扩展到了双类型情境。这种扩展的理论意义主要在于：

首先，加工分离法破除了经典两难法当中将行为反应与其背后的道德原则一一对应的关系。经典两难法中功利主义反应（即接受功利性的行为提议）与受功利主义原则驱动是对应的，而义务论反应（即拒绝功利性的行为提议）与受义务论原则驱动是对应的。加工分离法则破除了这种对应关系，人们在做出某个选择时，其背后的功利主义倾向与义务论倾向便都可以得到测量，而并非互斥关系。

其次，加工分离法引入了一致性两难情境作为参照，从而将功利主义原则和义务论原则都没有驱动行为反应的情形给分离出来，即图 1 当中最下面一行的情形。而这种情形在经典两难法当中一直被混淆在了功利主义倾向的表现当中。

Conway 和 Gawronski 在开发这一方法的同时，进行了三个实证研究，其结果进一步支持了道德双加工理论。他们发现，义务论倾向的确根植于对伤害行为的情绪反应，在共情关心和观点采择上的个体差异与反应义务论倾向的 D 参数相关，但与反应功利主义倾向的 U 参数不相关；强化共情关心也只会影响 D 参数而不影响 U 参数；认知需求的个体差异则与 U 参数相关，与 D 参数不相关，认知负荷的操控也只会选择性的影响了 U 参数而并不影响 D 参数。①

加工分离法对于在经典两难法研究中所得出的许多不一致的结论都给出了更多参考。比如，有研究使用经典两难法指出男性比女性更加功利主义，这在经典两难法视域下既可能是功利主义倾向更强也可能是义务论倾向更弱，而使用加工分离法则综合表明，女性比男性有中等程度的更高义务论倾向($d = 0.57$)，而男性比女性只有比较低程度的更高功利主义倾向($d = 0.10$)。这说明道德两难问题上的性别差异，主要是因为对伤害的情绪反应差异而不是对结果的认知评价差异的结果。②又如，有研究认为牺牲性判断所反应的是反社会性而并非真正的功利主义，③而使用加工分离法则反映出，反社会性预测义务论倾向的降低但不能预测功利主义倾向的升高，普通人牺牲性的功利主义判断也能反应其减少伤害的道德关注，无论是哲学家还是普通人，牺牲性功利主义判断都反映了真正的道德关怀。④

① P. Conway and B. Gawronski, “Deontological and utilitarian inclinations in moral decision making: a process dissociation approach,” *Journal of Personality and Social Psychology*, vol. 104, no. 2, 2013, pp. 216-235.

② R. Friesdorf, P. Conway and B. Gawronski, “Gender differences in responses to moral dilemmas: a process dissociation analysis,” *Personality and Social Psychology Bulletin*, vol. 41, no. 5, 2015, pp. 696-713.

③ G. Kahane, J. A. Everett, B. D. Earp, et al., “'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good,” *Cognition*, vol. 134, 2015, pp. 193-209.

④ P. Conway, J. Goldstein-Greenwood, D. Polacek, et al., “Sacrificial utilitarian judgments do reflect

chinaXiv:201911.00093v1

鉴于功利主义要求最大化结果，而义务论强调行为要符合道德规范，在研究中前者应对结果的利弊差异进行操控，而后者应对行为提议对道德规范的符合与否进行操控。因此，便形成了至少四种可能的组合情况：⑴行为提议的结果利大于弊，但为规范所禁止；⑵行为提议的结果弊大于利，且为规范所禁止；⑶行为提议的结果利大于弊，且为规范所提倡；⑷行为提议的结果弊大于利，但为规范所提倡。

尽管加工分离法在理论和方法上推进了道德困境研究，但其理论框架仍不完善。在规范（禁止或提倡）与结果（利大于弊或弊大于利）的四种可能组合中，只涵盖了其中两种，而缺乏对另外两种组合情形的考察。这一缺陷造成两个局限性：一是无法分离出决策者一般性不接受或者接受行为提议的倾向，在实际决策行为中存在着一种决策行为是并不关心决策背后的规范或结果，仅仅是一般性地对行为提议表示不接受或者接受，这种倾向无法在加工分离法中得到呈现；二是如同经典两难法将一般性接受倾向混淆在了功利主义倾向当中一样，加工分离法将一般性不接受倾向混淆在了义务论倾向当中，在图 1 的决策加工树第二行，决策者在一致和不一致情境下均选择不接受行为提议，这既可能是受义务论原则的驱动，也可能是一般性不接受行为提议而不顾规范是否合宜或结果是否有利。为了解决这一重要局限，Gawronski 研究团队于 2017 年使用计量经济学中的多项式决策加工树模型发展出了 CNI 模型法，既分离出决策受功利主义（或者结果主义）原则驱动的程度和受义务论（或者规范主义）原则驱动的程度，又分离出决策者不顾规范或结果而一般性不接受/接受倾向的程度。

（三）CNI 模型法

CNI 模型法在加工分离法基础上，将功利主义倾向操作化为对行为后果的敏感性，将义务论倾向操作化为对行为规范的敏感性，并进一步分离了一般性接受/不接受倾向，在理论上涵盖了规范（禁止或提倡）与结果（利大于弊或弊大于利）的四种组合，如表 1 所示。多项式决策加工树模型是 CNI 模型法的基础，在社会心理学的许多领域中具有广泛应用。①CNI 模型法能够同时量化出决策者的结果敏感性(Sensitivity to Consequences)，规范敏感性(Sensitivity to Norms)和（无视规范与结果的）一般性不接受或接受倾向(General preference for Inaction versus action irrespective of consequences and norms)，因此，该方法被称之为道德决策的 CNI 模型法。

表 1 CNI 模型法所使用情境示例②

情境	规范禁止		规范提倡	
	利大于弊	弊大于利	利大于弊	弊大于利
插队	你是一名普通旅客。在火车站售票大厅里，有个人需要购买一趟	你是一名普通旅客。在火车站售票大厅里，有个人需要购买一趟快	你是一名普通旅客。在火车站售票大厅里，有个人需要购买一趟快	你是一名普通旅客。在火车站售票大厅里，有个人需要购买一趟快要发车

concern for the greater good: Clarification via process dissociation and the judgments of philosophers," *Cognition*, vol. 179, 2018, pp. 241-265.

① M. Hutter and K. C. Klauer, "Applying processing trees in social psychology," *European Review of Social Psychology*, vol. 27, no. 1, 2016, pp. 116-159.

② C. Liu and J. Liao, "The Motivation Process of the Moral-question Framing Effect is Explained by the Action-based Model of Cognitive Dissonance," *Cognition*, in review, for a preprint: http://www.chinaxiv.org/businessFile/201904/.00087v4/.00087v4.pdf.

chinaXiv:201911.00093v1

情境	快要发车的火车票而且当天仅此一班。帮助他插队买票乘车的话，他一行五人可以顺利出行，但会影响另一个人出行的时间。	要发车的火车票。帮助他插队买票的话，他能顺利买到票乘车,但会导致排队的五个人无法买到当天仅有一班的车票而无法正常出行。	要发车的火车票，正在插队买票。阻止他插队的话，他将不能按时出行，但可以保证排队的大约五个人可以顺利买到当天仅有一班的车票。	的火车票而且当天仅此一班。他正在插队买票，阻止他插队的话，他将不能按时出行，并且阻止过程引起的混乱会导致排队的五个人也无法按时买到票出行。
	你打算帮他插队。	**你打算帮他插队。**	**你打算阻止他插队。**	**你打算阻止他插队。**

CNI 模型法假定了决策者在进行道德决策时，遵循序列加工规律，首先考虑行为提议的结果是否有利，再考虑行为提议是否符合规范，两者均不考虑的情况下，则会选择一般性地接受或者不接受行为提议。因此，使用决策加工树模型可以将这些不同倾向逐层剥离出来，如图 2 所示。当受功利主义原则驱动时，决策者在结果有利时接受提议而在结果有弊时不接受提议（图 2 第一行）；如果没有受功利主义原则驱动而受义务论原则驱动，决策者便会在规范提倡时接受提议而在规范禁止时不接受提议（图 2 第二行）；如果既没有受功利主义原则驱动，也没有受义务论原则驱动，决策者便会选择一般性地不接受（图 2 第三行）或者接受（图 2 第四行）行为提议。

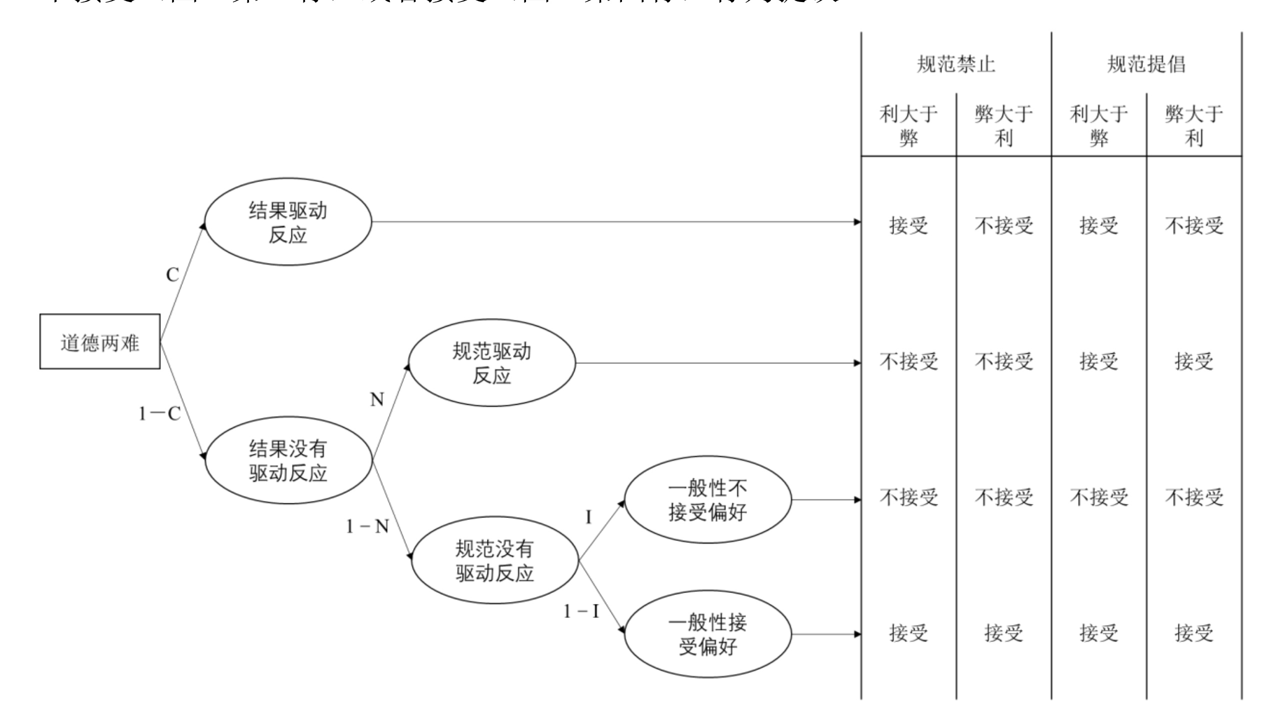


图 2 CNI 模型的多项式决策加工树图解，其中 C 代表结果敏感性，N 代表规范敏感性，I 代表一般性不接受偏好（一般性接受偏好的反应概率与一般性不接受偏好的反应概率之和为 1，因此在模型中仅用一个 I 参数即可表达）①，刘嫒嫒等人也曾对此模型进行了译介。②

在 CNI 模型法中，决策者在不同条件下的决策结果所反映出的概率原理与加工分离

① B. Gawronski, J. Armstrong, P. Conway, et al., “Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making,” *Journal of Personality and Social Psychology*, vol. 113, no. 3, 2017, pp. 343-376.

② 刘嫒嫒,丁一,彭凯平,胡传鹏：《多项式加工树模型在社会心理学中的应用》,《心理科学》2019 年第 42 卷第 2 期.

chinaXiv:201911.00093v1

法类似：

p(接受|规范禁止，利大于弊) = C + (1－C) × (1－N) × (1－I)　(5)

p(接受|规范禁止，弊大于利) = (1－C) × (1－N) × (1－I)　(6)

p(接受|规范提倡，利大于弊) = C + (1－C) ×N + (1－C) × (1－N) × (1－I)　(7)

p(接受|规范提倡，弊大于利) = (1－C) ×N + (1－C) × (1－N) × (1－I)　(8)

在同类情境之下的多个情境故事之后，选择接受和不接受的概率之和为 1，比如，在规范禁止而利大于弊的情况下，有 6 个情境故事，决策者对其中 2 个故事的行为提议选择了接受，则其接受概率为 1/3，不接受概率即为 2/3。因此，同类情境下接受与不接受反应的概率等式在统计上具有等价意义，此处只列了在四种组合情境下决策者接受提议的概率等式，为了简化表达式，依次使 p(接受|规范禁止，利大于弊)为 $p1$，p(接受|规范禁止，弊大于利)为 $p2$，p(接受|规范提倡，利大于弊)为 $p3$，p(接受|规范提倡，弊大于利)为 $p4$，下文相同。

为了最小化经验观察到的决策者在四种情境中接受/不接受反应的可能性，与使用指定参数估计的模型等式所预测的接受/不接受反应的可能性之间的差异，多项式决策加工树模型使用了最大似然估计的方法来实现这一目标。模型对数据刻画的准确性可以通过平均拟合优度(Means of goodness-of-fit statistics)来检验，如果模型适配较差就说明经验观察到的可能性与模型预估的可能性之间变异较大，违背了模型的基本假设。模型适配变异程度可以通过计算 Cohen’s w 来表示，0.1 表示变异较少，0.3 表示变异中等，0.5 表示变异较大，①但如果抽样样本量很大时，模型适配变异程度可以不作考察，因为样本量的增大本身也会导致模型适配变异度的增加。

通过 Gawronski 提供的决策加工树工具可以直接计算出决策者的结果敏感性（C 参数）、规范敏感性（N 参数）和一般性不接受/接受倾向（I 参数）(http://www.bertramgawronski.com/documents/CNI-Model_Materials.zip)，同时，也可以对这三个参数在两组数据之间进行比较或者与某个特定值进行比较。所得到的 C 和 N 参数如果显著大于 0，则表明决策者有显著的结果和规范敏感性，如果 I 参数显著大于 0.5 则表明决策者有显著的一般性不接受倾向，如果显著小于 0.5 则表明决策者有显著的一般性接受倾向。

前已述及，为了解决经典两难和加工分离法所存在的局限性，Gawronski 团队发展出了 CNI 模型法。②③这一方法对于解决道德困境研究中的争议问题也具有促进作用。比如，前人发现情绪对于道德判断具有重要作用，但不能确定情绪作用于规范敏感性还是结果敏感性，抑或只是作用于一般性反应倾向，而使用 CNI 模型法则表明，幸福感减少了对道德规范的敏感性而没有影响对结果的敏感性或一般性不接受倾向；悲伤和愤怒

① J. Cohen, *Statistical power analysis for the behavioral sciences*, Hillsdale, NJ: Erlbaum, 1988.

② B. Gawronski and J. S. Beer, “What makes moral dilemma judgments "utilitarian" or "deontological"?” *Social Neuroscience*, vol. 12, no. 6, 2017, pp. 626-632.

③ B. Gawronski, J. Armstrong, P. Conway, et al., “Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making,” *Journal of Personality and Social Psychology*, vol. 113, no. 3, 2017, pp. 343-376.

chinaXiv:201911.00093v1

对道德两难判断没有显著影响。①前人有研究表明睾丸酮素会增强决策者的功利主义反应倾向，而使用 CNI 模型法则发现，外源性类固醇睾丸酮素使决策者在规范禁止而利大于弊的情境下不接受行为提议的倾向增强，这种反应是由于对规范的敏感性增强导致的；相反，内源性睾酮在基线测量时的模式正好相反，内源性睾酮水平越高，对道德规范的敏感度越低。研究结果表明，睾丸激素在道德判断中的作用比之前的研究结果更为复杂。②这些研究均展现了 CNI 模型这样的形式化建模方法对于深入探讨道德判断决定因素的方法价值。

CNI 模型法在理论构念上全面覆盖了规范（禁止或提倡）与结果（利大于弊或弊大于利）的四种组合情况，在方法论上具有突破性的贡献。但是，这一方法依然存在着一些局限性：

首先，该方法不能应用于相关或回归研究设计。CNI 模型法的底层是多项式决策加工树模型，这种方法广泛应用于区分前置变量中的多重影响因素。但是，这种方法是以群体比较而非个体比较为基础的，所生成的 C、N、I 参数是表征在群体层面而非个体层面的，因此不同进行相关或回归分析。

其次，受限于多项式加工决策树工具，也只能比较两组参数之间或者与参照值之间的差异性检验，而不能进行多组数据之间的比较。因此，其计算方法的拓展性较为受限。

再次，CNI 模型法无法进行简单效应分析，不能更深入分析跨组数据之间的差异究竟存在于四种结构情境中的哪一种。比如，Gawronski 等(2017)发现在行动框架问题下比判断框架问题下，决策者的规范敏感性会下降，一般性不接受倾向会升高，然而，对于这种框架效应究竟存在于四种结构情境中的哪一种，CNI 模型法不能给出更多解答。

最后，CNI 模型法最大的问题在于其理论逻辑先验地假定了决策者的决策过程是一种特定的序列加工过程：决策者首先考虑的是结果，在不考虑结果的情况下再考虑规范，在结果与规范均不考虑的情况下，才会表现出一般性的不接受或者接受倾向。CNI 模型法的构念逻辑是逐层剥离，而非平行建构的。而实际上，决策者完全有可能是并行加工进行决策的，在决策时同时考虑决策背后的规范以及决策可能导致的结果，抑或先考虑决策是否符合规范然后再考虑决策的结果，也有可能首先只是形成了一般性行为态度，然后受到规范或结果原则的修正。因此，倘若将图 2 当中各参数的位置互换之后，其算法结果的概率模型将完全不同，下文详述。

三、新开发范式——CAN 维度合成法

前已述及，CNI 模型法在原来的经典两难法和加工分离法基础上，在理论和方法上均具有突破性的贡献。然而，CNI 模型法的理论和算法也存在局限。特别是在理论构念上，它假定了决策者首先考虑行为结果，然后再考虑行为的规范性，最后才考虑一般性行为倾向的梯级序列。这一在构念上的基本假设尚无研究对其进行检验。

如果 CNI 模型法所预设的决策者遵循序列加工是符合实际情况的，那么，决策者除了按 CNI 模型法所假定的结果→规范→一般性不接受/接受倾向序列进行决策之外，

① B. Gawronski, P. Conway, J. Armstrong, et al., “Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model,” *Emotion*, vol. 18, no. 7, 2018, pp. 989-1008.
② S. M. Brannon, S. Carr, E. S. Jin, et al., “Exogenous testosterone increases sensitivity to moral norms in moral dilemma judgements,” *Nature Human Behaviour*, vol. 3, no. 8, 2019, pp. 856-866.

chinaXiv:201911.00093v1

也可能按规范→结果→一般性不接受/接受倾向，或者一般性不接受/接受倾向→受规范修正→受结果修正，或者一般性不接受/接受倾向→受结果修正→受规范修正等序列加工思路进行决策。试以规范→结果→一般性不接受/接受倾向这一序列为例，当决策者按这一序列进行决策加工时，其决策树模型如图 3 所示：

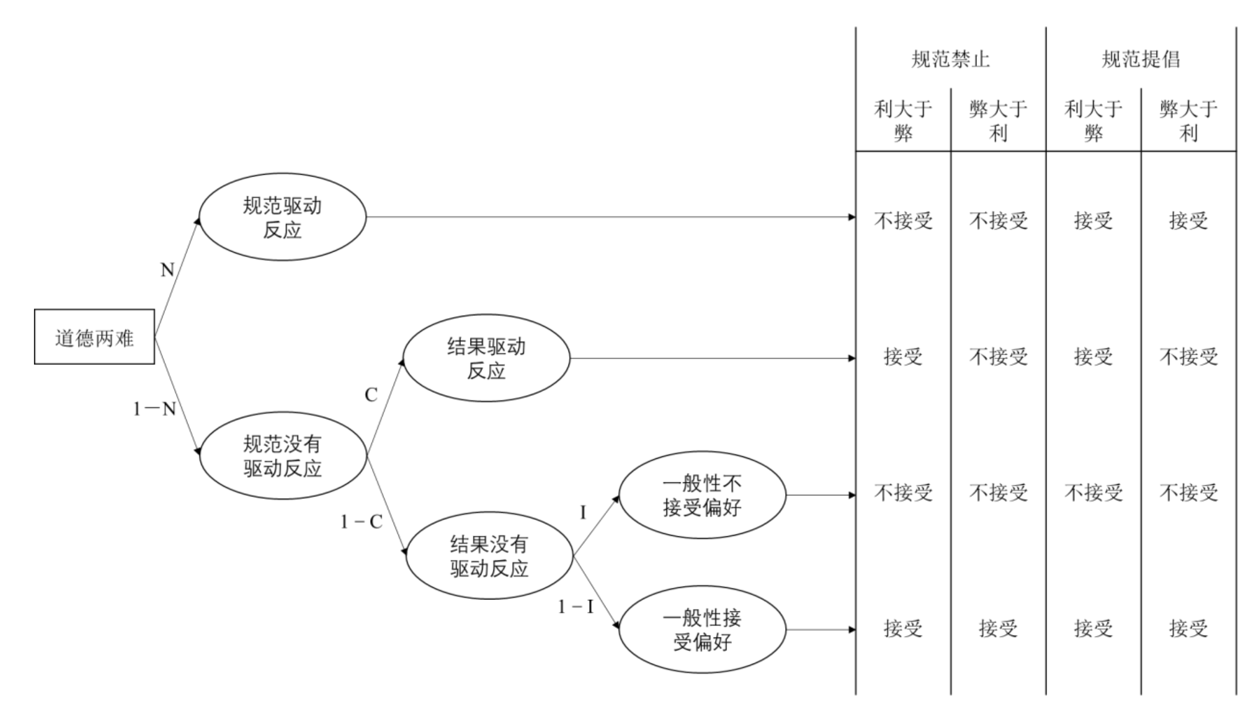


图 3 规范→结果→一般性不接受/接受倾向下的决策加工树模型

相应地，决策者在各类情境下的决策反应概率为：

$$p1 = (1-N) \times C + (1-N) \times (1-C) \times (1-I) \quad (9)$$

$$p2 = (1-N) \times (1-C) \times (1-I) \quad (10)$$

$$p3 = N + (1-N) \times C + (1-N) \times (1-C) \times (1-I) \quad (11)$$

$$p4 = N + (1-N) \times (1-C) \times (1-I) \quad (12)$$

倘若按照 CNI 模型法所假定的，决策者在决策加工序列上对其决策倾向的估计是没有影响的，那么，等式(5)~(8)与等式(9)~(12)在统计上应当是分别等价对应的，但是稍加换算便可发现这种情形在统计上成立的可能性极小，比如将图 2 模型的概率等式中 N 参数经换算后得到 $N=(-p1-p2+p3+p4) / (2-p1+p2-p3+p4)$；而将图 3 模型的概率等式中 N 参数经换算后得到 $N=(-p1-p2+p3+p4) / 2$。如果 CNI 模型中的参数位置关系并不会影响概率估计结果的话，这两个 N 参数应当相等。如此，则会换算出 $p2-p1 = p3-p4$，这在经验统计上成立的概率极小。Gawronski 等(2017)在脚注 7 里报告称，当 C 和 N 的位置互换之后，可以重复出他们研究的效应结果，但对于部分差异边缘显著的结果就会变得显著，因此，作者未对此更进一步分析，也未说明为何会出现统计变得更显著的原因。①但是，这恰恰也说明 C 和 N 的位置对于最终的参数估计结果存在影响，只是这种影响恰好符合作者的假设便未得到进一步辨析。

① B. Gawronski, J. Armstrong, P. Conway, et al., "Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making," *Journal of Personality and Social Psychology*, vol. 113, no. 3, 2017, pp. 343-376.

chinaXiv:201911.00093v1

此外，Gawronski 等(2017)认为一般性行为反应倾向只能在模型的最低层级，因为他们的反应概率之和为 1，是非此即彼的矛盾反应倾向。这一先验假设也未必成立。决策者完全可能首先拥有一般性接受或不接受反应倾向，而后受到规范或结果原则的修正而改变了原来的反应倾向，即在图 2 和图 3 当中 C，N，I 三个参数之间的位置关系在逻辑上是不确定和可以互换的。

从以上分析可以得出，决策者的决策模式如果是序列加工形态的话，应当存在着多种决策模式，而 CNI 的模型算法显然只是针对其中一种，这一种在经验上的确有可能是多数决策者所具有的决策模式。但是，我们并不能忽略其他决策模式存在的可能性，并且，决策者的确更加可能会同时对规范与结果进行权衡，才会出现冲突性的两难。如果决策者只是序列性地选择了其中某一条决策路径，则在实际决策时不会出现规范与结果相冲突的情形。因此，在存在潜在两难情境问题中，决策加工过程更可能是并行而非序列加工形态。

为了解决上述局限性，结合 CNI 模型的理论构念，实际上在道德决策中应综合考虑决策者在 2（规范：提倡/禁止）×2（结果：利大于弊或弊大于利）的四种组合情况下，决策者的不同反应倾向。比如在 CNI 模型的情境结构中，在上述四种组合情况下均有六个情境故事，因此，可以从经验概率上计算出每种组合情况下个体在六个情境故事中选择接受行为提议的概率。这四个概率数据在研究设计上形成了被试内 2 × 2 的结构，可以将规范与结果当作两个平行的被试内因子作为方差设计的组内因子，如果要做组间比较，可以将分组变量作为组间因子。如果规范或结果因子具有主效应，说明决策者受到了规范或结果原则的驱动（即规范提倡或结果利大于弊时接受提议的概率显著大于规范禁止或结果弊大于利时），或者反规范或反结果原则的驱动（即规范提倡或结果利大于弊时接受提议的概率显著小于规范禁止或结果弊大于利时）。如果规范与结果的交互作用显著，说明两种原则的确存在交互影响，而进一步简单效应分析则可以找出二者具体的交互模式。同理，组间效应与组间组内交互效应及其简单效应，均可以得到解答，值得注意的是由于进行计算的是决策反应的概率数据，在进行多重比较时，推荐使用 Bonferroni 检验。

在方差分析的基础上，综合考虑决策者可能的并行加工的决策特点，也可以合成与 CNI 模型类似的三个维度指标，①为了与 CNI 模型法作为区分，并在提取维度指标首字母时考虑方便记忆，称之为 CAN 维度合成法：

结果敏感性指标 C(Consequence sensitivity, C) = $1/2(p1-p2+p3-p4)$；

规范敏感性指标 N(Norm sensitivity, N) = $1/2(p3-p1+p4-p2)$；

一般性接受/不接受指标 A(generalized Action preference, A) = $1/4(p1+p2+p3+p4)$。

CAN 维度合成法从并行加工角度，假定了决策者同时考虑到规范与结果两个方面，

① 这种合成方法也常见于文献当中，比如 Talhelm 计算忠诚/裙带关系(loyalty/nepotism)指标时，用参与者奖励朋友的数量减去惩罚朋友的数量(T. Talhelm, X. Zhang, S. Oishi, et al., “Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture,” *Science*, vol. 344, no. 6184, 2014, pp. 603-608.)，与本研究中使用规范提倡时对行为提议的赞成概率减去规范禁止时对行为提议的赞成概率来表示规范驱动指标、使用结果利大于弊时对行为提议的接受概率减去结果弊大于利时对行为提议的接受概率来表示结果驱动指标，在原理上是相通的。

chinaXiv:201911.00093v1

并且受二者交互影响。如果规范对决策者具有正面影响，那么，无论结果利大于弊或者弊大于利时，规范提倡情况下决策者对行为提议的接受概率应当显著大于规范禁止时的接受概率，二者之差即可代表规范的敏感程度。当利大于弊时，规范的敏感程度可以用 $p3-p1$ 来表示，当弊大于利时，规范的敏感程度可以用 $p4-p2$ 来表示。那么，对规范的整体敏感程度也就可以用二者的平均数来表示，即 $1/2(p3-p1+p4-p2)$。同理，如果结果对决策者具有正面影响，那么，无论规范提倡或者禁止时，结果利大于弊的情况下决策者对行为提议的接受概率应当显著大于弊大于利时的接受概率，二者之差即可代表结果的敏感程度。当规范禁止时，结果的敏感程度可以用 $p1-p2$ 来表示，当规范提倡时，结果的敏感程度可以用 $p3-p4$ 来表示。那么结果的整体敏感程度也就可以用二者的平均数来表示，即 $1/2(p1-p2+p3-p4)$。而作为整体反应倾向指标而言，所有情形下的平均接受概率则可以反映决策者的整体偏好性，即 $1/4(p1+p2+p3+p4)$。

例如，假设某个体在规范与结果的四种组合情况下，多个情境故事中选择接受行为提议的概率分别为 $p1=0.6, p2=0.1, p3=0.9, p4=0.5$。那么，其规范敏感度 $1/2(p3-p1+p4-p2)=0.35$；其结果敏感性 $1/2(p1-p2+p3-p4)=0.45$；其整体偏好性 $1/4(p1+p2+p3+p4)=0.525$。同理，即可计算出每一个个体在这三个指标上的值，然后进行统计检验。

在 CAN 维度合成法的指标解读上，C 指标如果在统计上显著大于 0，说明决策者在行为提议造成结果利大于弊的情况下对提议的接受概率高于在行为提议造成结果弊大于利的情况下对提议的接受概率，因此受到显著的功利主义（或者结果主义）原则驱动，如果显著小于 0，说明决策者受到显著的反功利主义（或者反结果主义）原则驱动，如果与 0 无显著差异，说明决策未受到功利主义原则的驱动；同理，N 指标如果显著大于 0，说明决策者受到显著的义务论（或者规范主义）原则驱动，如果显著小于 0，说明决策者受到显著的反义务论（或者反规范主义）原则驱动，如果与 0 无显著差异，说明决策者未受到义务论原则的驱动。对于 A 指标，如果显著大于 0.5，说明其具有一般性接受倾向相比于不接受倾向的优势，如果显著小于 0.5，说明其具有一般性不接受倾向相比于接受倾向的优势，如果与 0.5 无显著差异时有两种情况：1、如果 C 指标和 N 指标同时与 0 也无显著差异，说明决策者只是完全随机作答；2、如果至少 C 和 N 指标其中之一与 0 有显著差异，说明决策者总体上的一般性接受与不接受倾向强度相当，但也受到了功利主义或义务论原则的影响。

CAN 维度合成法是建立在 CNI 模型法基础上的，全面汲取了 CNI 模型法的理论构念优势，但二者在五个方面存在着差异：第一，CNI 模型法假定了决策者在进行道德决策时是基于先结果后规范再一般性不接受/接受偏好的序列加工模式，而 CAN 维度合成法则没有这样的先验假设，而使用了一个常见的参数合成中的减法策略，将规范与结果的作用同等看待，取其平均。第二，CNI 模型法导出的参数是在群体层面的，因此不能用于相关或回归设计，参数也只能在两组之间比较或者与某一特定值进行比较，而 CAN 维度合成法所得到的参数是在个体层面的，既可以用于相关或回归设计，也可以在多组之间进行比较，还可以在方差分析方法下探讨其简单效应。第三，CNI 模型法中 I 因子与 CAN 维度合成法中的 A 因子在统计上的解释方向是相反的，前者数值越大代表一般性不接受倾向越强，后者数值越大代表整体性接受倾向越强。第四，CNI 模型中的 I 因

chinaXiv:201911.00093v1

子是在序列加工为假设前提下剥离了结果和规范的作用之后，代表决策者一般性不接受/接受倾向的，而 CAN 维度合成法中的 A 因子是没有剥离结果或规范作用的前提下，代表决策者的整体上接受/不接受倾向，在内涵上不同。第五，CNI 模型法依赖于二元反应模式，个体必须对行为提议做二元判断，要么接受要么不接受，而 CAN 维度合成法无此要求，也可用于连续性评分设计当中，决策者通过对行为提议的接受或不接受程度进行评分也可以实现相应的参数计算。

综合来看，CNI 模型法和 CAN 维度合成法，在实际研究中可以综合运用，相互参照。在本文研究团队中，已综合使用上述四种研究方法来探讨道德问题框架效应，并证明了 CAN 维度合成法的统计检验结果的有效性，但具有可探索简单效应，可用于相关与回归设计，可进行多重比较等优势。①

四、讨论与展望

道德困境研究是广泛的社会困境研究中的典型代表，其研究方法对于其他社会困境研究具有强大的参考性和可迁移性。纵览道德困境研究的四种实证方法，从经典两难法和加工分离法，到 CNI 模型法和本研究提出的 CAN 维度合成法，在理论构念上是向下兼容的，越往后的方法涵盖了之前方法的理论构念。特别是 CNI 模型法和 CAN 维度合成法，实现了对规范（禁止或提倡）与结果（利大于弊或弊大于利）的全组合情境的通盘考察，在未来的道德研究当中应该重点加以应用，在其他社会困境研究中也可以迁移使用。

CNI 模型法和 CAN 维度合成法对于以往悬而未决的许多争议均具有很强的指导意义。首先，以往研究中所发现的精神病性与功利主义反应的正相关关系，可能只是因为精神病性与一般性不接受/接受倾向存在相关，而与结果敏感性未必存在着关联，这需要后续研究进一步检验。其次，前人研究表明厌恶感会使道德判断更为严厉，但近期研究表明并无此效应，②此外，情绪性厌恶唤醒究竟是强化了个体的规范敏感性，还是只是增强了其一般性不接受倾向，也可以通过本研究中的后两种方法得到解答。再次，个体认知的双加工状态与功利主义/义务论倾向之间是否具有一一对应关系，也需要重新检验，有学者提出存在直觉功利主义的可能，③④这也可以使用后两种方法来进行检验。更进一步，近来有学者使用 CNI 模型的变体来考察规范/结果敏感性与义务论/功利主义的对应关系，发现代表规范的参数对结果也很敏感，由此得出，除非期望规范产生切实的结果，否则它们不会指导道德判断，这表明规范和结果(或义务论和功利主义)作为判

① C. Liu and J. Liao, “The Motivation Process of the Moral-question Framing Effect is Explained by the Action-based Model of Cognitive Dissonance,” *Cognition*, in review, for a preprint: http://www.chinaxiv.org/businessFile/201904/.00087v4/.00087v4.pdf.

② D. J. Johnson, J. Wortman, F. Cheung, M. Hein, R. E. Lucas, M. B. Donnellan, . . . R. K. Narr, “The Effects of Disgust on Moral Judgments,” *Social Psychological and Personality Science*, vol. 7, no. 7, 2016, pp. 640-647.

③ B. Bago and W. De Neys, “Advancing the specification of dual process models of higher cognition: a critical test of the hybrid model view,” *Thinking & Reasoning*, 2019, DOI: 10.1080/13546783.2018.1552194.

④ B. Bago and W. De Neys, “The intuitive greater good: Testing the corrective dual process model of moral cognition,” *Journal of Experimental Psychology: General*, vol. 148, no. 10, 2019, pp. 1782-1801.

chinaXiv:201911.00093v1

断的决定因素之间的分裂是人为的，①而实际上这也就是 CAN 维度合成法当中，存在着规范与结果的交互作用。其他的许多研究争议也可以在本研究框架下得到更深入解答。

CNI 模型法和 CAN 维度合成法还具有非常强大的理论和方法拓展性。这种方法作为一种理论和方法思路，并不仅仅只能用于道德困境的研究，也可以应用于其他具有潜在冲突性的研究议题上，试举三例。在行为的道德可责性确定上一直存在着意图与后果的争议，意图论者认为行为的道德可责性在于其意图是否存在主观故意，而后果论者认为行为的道德可责性在于其后果是否存在伤害或妨碍性。那么，对于行为的道德可责性的确定，意图与后果便形成了潜在的冲突，可能存在故意而有害、故意而无害、非故意而有害和非故意而无害等情境组合，可以结合 CNI 模型法和 CAN 维度合成法的理论构念对此进行研究。近期有学者在研究对伤害的情绪反应时，也推荐了这种进一步检验思路。②同理，在个人道德偏好与道德抉择上，中国传统文化中一直存在着公义与私利的争论，公义与私利常常形成潜在冲突，出现益公益私、益公害私、害公益私和害公害私等情境组合，也可以通过类似的方法进行剖析。再如，在消费决策当中，性能与价格之间的性价比问题也是一对常见矛盾，可能存在高性能高价格、高性能低价格、低性能高价格、低性能低价格等不同组合情形，同样可以使用上述方法进行深入挖掘。因此，CNI 模型法和 CAN 维度合成法，除了推进道德困境的研究之外，具有跨学科跨领域应用性，只要具有潜在冲突性的研究议题都可以考虑使用与之类似的方法论框架来进行探讨。

综上，道德困境研究走过了经典两难法、加工分离法、CNI 模型法和 CAN 维度合成法四种方法学沿革，在理论构念上从单类型冲突情境，向规范（禁止或提倡）与结果（利大于弊或弊大于利）的四种全组合情境发展。特别是 CNI 模型法和 CAN 维度合成法，对于解决以往研究中的许多争议提供了方法学思路。并且，这种思路框架并不局限于道德困境研究，也为跨专业跨领域中具有潜在冲突性的社会困境和其他议题提供了方法学借鉴。

chinaXiv:201911.00093v1

① M. Hennig and M. Hütter, “Revisiting the Divide between Deontology and Utilitarianism in Moral Dilemma Judgment: A Multinomial Modeling Approach,” *Journal of Personality and Social Psychology*, 2019, DOI: 10.1037/pspa0000173.

② C. J. Reynolds and P. Conway, “Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas,” *Emotion*, vol. 18, no. 7, 2018, pp.1009-1023.

Methodological Paradigms and Their Theoretical Values of Social Dilemma Research

Liu Chuanjun and Liao Jiangqun

（*Department of Psychology, School of Social Sciences, Tsinghua University, Beijing 100084*）

ABSTRACT

Social dilemma is a multi-disciplinary topic widely discussed in Ethics, Psychology, Sociology and Economics, among which moral dilemma is the typical paradigm and has a long history of research. Present article reviewed three existed empirical approaches in dilemma research, traditional dilemma paradigm, process dissociation method and Consequences-Norms-generalized Inaction/action (CNI) model. Based on analyzing the contributions and limitations of the above three approaches, we developed a Consequences-general Action/inaction preference-Norms sensitivity estimation (CAN) algorithm. With the development of the four approaches, moral dilemma research has no longer limited in the contradicted dilemma situations. CNI model and CAN algorithm have extended to consider the four combinations between proscriptive/prescriptive norms and benefits greater/smaller than costs. With CNI model and CAN algorithm, Controversies in moral theories and empirical inconsistences can be further clarified. The four approaches provided methodological references for the similar topics with potential contradictions in many other domains. Therefore, the approaches can be used in multi-disciplines.

Key words: empirical approaches; social dilemma; moral dilemma; conflict rules

chinaXiv:201911.00093v1

作者简介：
刘传军，清华大学心理学系社会心理学在读博士生，中级职称，联系电话：13550836591，电子邮箱：lcj17@mails.tsinghua.edu.cn， 通信地址：北京市海淀区清华园 1 号清华大学社会科学学院心理学系伟清楼，邮编 100084。
廖江群，博士，清华大学心理学系副教授，博士生导师，联系电话：010-62787208，电子邮箱：liaojq@tsinghua.edu.cn， 通信地址:北京市海淀区清华园 1 号清华大学社会科学学院明斋 210 室，邮编 100084。

chinaXiv:201911.00093v1